Maternal Health Risk Classification

Lily Vogel

DS 480 Data Science Capstone

Quinnipiac University

Spring 2025

Abstract

Despite medical advances, the identification of maternal health risk level is a current worldwide issue. This study aimed to improve the classification of maternal health risk using machine learning models. I used a dataset containing maternal health features with their associated risk level. Using ordinal probit logistic regression, a decision tree, and a random forest, I identified the importance of each feature in predicting maternal health risk as well as which model performed the best for the classification task based on overall accuracy and within-risk performance. In terms of feature importance, my results showed that the most important predictor for classifying maternal risk level is blood glucose level in all three models. In terms of model comparison, the random forest model outperformed the other two models in terms of accuracy and within-risk performance. Overall, this study contributed to the evaluation of machine learning methods in improving the assessment of maternal health risk.

Introduction

Advancements in machine learning (ML) methods within the healthcare system have led to earlier detection, diagnosis, and treatment strategies (Goecks et al., 2020). Machine learning models can recognize complex patterns within medical data that could be overlooked by humans or traditional statistical analyses (Sathasivam and Abdullahi, 2024). Research from numerous studies have found that ML algorithms can uncover insights from large datasets that medical professionals could not identify themselves (e.g., Zhang et al. (2019); Tomczak et al. (2015); Venter (2001)). An application of ML models in healthcare is risk assessment, where models use data to classify levels of health risk across various conditions. For instance, ML techniques have been used to assess the likelihood of mild cognitive impairment by classifying the risk level of individuals (Liu et al., 2025). Another important example of this application is the classification of maternal health risk. Maternal health risk refers to the likelihood that a mother or child experiences harmful outcomes during the pregnancy, labor, or postpartum period (Togunwa et al., 2023).

Unfortunately, maternal health risk remains a global concern. Although many researchers have investigated ways to reduce the risk of maternal health, maternal mortality rate has not consistently declined across all countries in the world (Ahmed et al., 2020). According to the World Health Organization, nearly 800 women died every day on average from preventable issues during pregnancy and childbirth in 2020 (World Health Organization, 2024). The ability to identify risks can significantly improve the outcome of pregnancy experiences and even lower the maternal mortality rate (Macrohon et al., 2022). While previous studies have explored machine learning techniques for maternal health risk classification, there are still limitations in model variation and feature importance analysis. For example, Sathasivam and Abdullahi (2024) conducted a previous study in which they used a machine technique to create a hybrid model consisting of the combination of support vector machines and logistic regression to classify maternal health risk level. In another study, Mondal et al. (2023) collected data from Internet of Things (IoT) devices to predict high risk pregnancy complications. Their study is limited in terms of feature importance analysis since their IoT devices only collected data on three features (heart rate, blood pressure, and body temperature). Given these gaps in research, this study aimed to improve these issues by examining multiple ML methods to assess maternal health risk using a dataset with more maternal health features to identify the most crucial ones.

In this study, I aimed to improve maternal health risk classification by addressing two research goals using a dataset with maternal health features. The dataset includes six maternal health features: age, systolic blood pressure, diastolic blood pressure, blood glucose level, body temperature, and heart rate. First, I used three machine learning models—an ordered probit logistic regression, a decision tree, and a random forest—to identify the most important factors for classifying maternal health risk. Second, I evaluated and compared overall model accuracy as well as within-group performance using confusion matrices. By understanding key risk factors and improving accuracy, this study contributes to the effort to use machine learning to improve maternal healthcare outcomes. These goals are essential for developing effective maternal health risk assessments that can assist healthcare professionals in making informed decisions.

Methods

Dataset

All data came from the Maternal Health Risk dataset from the University of California, Irvine (UCI) Machine Learning Repository (Ahmed, 2020). The Maternal Health Risk dataset contained data collected from hospitals, community clinics, and other healthcare facilities from Bangladesh. The dataset contained six features, including age, systolic blood pressure, diastolic blood pressure, blood glucose levels, body temperature and heart rate. The target variable, risk level during pregnancy, had three different classes. The three classes were high risk, mid risk, and low risk. In total, the dataset consisted of 1,013 observations in which 406 were categorized as low risk, 336 were mid risk, and 272 were high risk.

Predictors of Maternal Health Risk

To identify the most important predictors of health risk during pregnancy, I fit an ordinal logistic regression, decision tree, and random forest model using tools from the statsmodels and scikit-learn libraries (Seabold and Perktold, 2010; Pedregosa et al., 2011). I found the optimal distribution for the ordinal logistic regression model by comparing the logit and probit models. I fit two ordinal logistic regression to the training data using each distribution and evaluated their individual AIC scores. The probit model had a lower AIC and thus was chosen as the better fit. The decision tree and random forest models were fit with optimized hyperparameters found in k-fold cross-validation to estimate feature importance. I optimized the the maximum depth in my decision tree model and the maximum number of features per split in my random forest model. For the ordinal probit logistic regression model, feature importance was assessed based on the magnitude of coefficients. For the decision tree model, feature importance was accessed based on the split criteria of the initial nodes and the normalized sum of impurity decrease that each feature contributes across nodes where it was used. For the random forest model, feature importance was computed as the mean and standard deviation of the accumulation of the impurity decrease in each tree.

Model Comparison

To evaluate model accuracy in classifying health risk during pregnancy, I compared the accuracy of the ordinal logistic regression, decision tree, and random forest models. Accuracy was calculated by the sum of correct classifications divided by the total number of predictions. Additionally, I looked at within-risk category accuracies by creating a confusion matrix for each model.

Results

Predictors of Maternal Health Risk

Table 1 presents the predictor variables included in the ordinal probit logistic regression model, along with the absolute values of their corresponding coefficients. The features were ranked in descending order based on their coefficient magnitudes, indicating their relative influence on the model's predictions. Blood glucose level was the most important feature for classification, with an importance score of 0.822. Systolic blood pressure and body temperature followed, with scores of 0.459 and 0.417, respectively. The remaining features—heart rate, diastolic blood pressure, and age—had lower importance values, with age ranked last at 0.001.

Figure 1 presents the decision tree model, fitted with an optimal maximum depth of 4, used to predict maternal risk level. The feature appearing in the root node has the highest importance, as it contributed significantly to the model's decision-making process. Features closer to the root node had higher importance than those that are further away. Features

Table 1: The predictor variables present in the ordinal logistic regression model with the absolute value of their corresponding coefficients, sorted in descending order.

Rank	**Feature**	**Coefficient**
1	Blood Glucose Level	0.821944
2	Systolic Blood Pressure	0.458814
3	Body Temperature	0.417068
4	Heart Rate	0.213583
5	Diastolic Blood Pressure	0.026549
6	Age	0.000755

that were not included in the decision tree were considered low importance since they did not contribute to the decision-making process. Blood glucose level (BS) appeared as the root node and appeared frequently throughout the tree, indicating its strong role in splits. Systolic blood pressure also appeared in multiple nodes, often near the top of the tree, suggesting it was another key feature in the decision-making process. Other features such as body temperature, diastolic blood pressure, and age appeared in lower levels of the tree. Heart rate did not appear in any of the decision nodes, consistent with its feature importance score of 0 in Table 2.

The class predictions at the leaf nodes show a mix of all three risk levels (low, mid, high), with some branches ending in pure class nodes (e.g., [0, 0, 68]) and others containing mixed class distributions.

Table 2 presents the predictor variables included in the decision tree model, along with their corresponding importance. Importance was computed as the normalized sum of total impurity decrease. The features were ranked in descending order based on their importance within the table. Matching the patterns seen in Figure 1, blood glucose level was the most important feature for classification, with an importance score of 0.635. Systolic blood pressure and body temperature followed, with scores of 0.028 and 0.04, respectively. The remaining features—age, diastolic blood pressure, and heart rate—had lower importance values. Heart rate ranked last at 0, indicating that it was not used in any of the splits made by the decision tree like we saw in Figure 1.

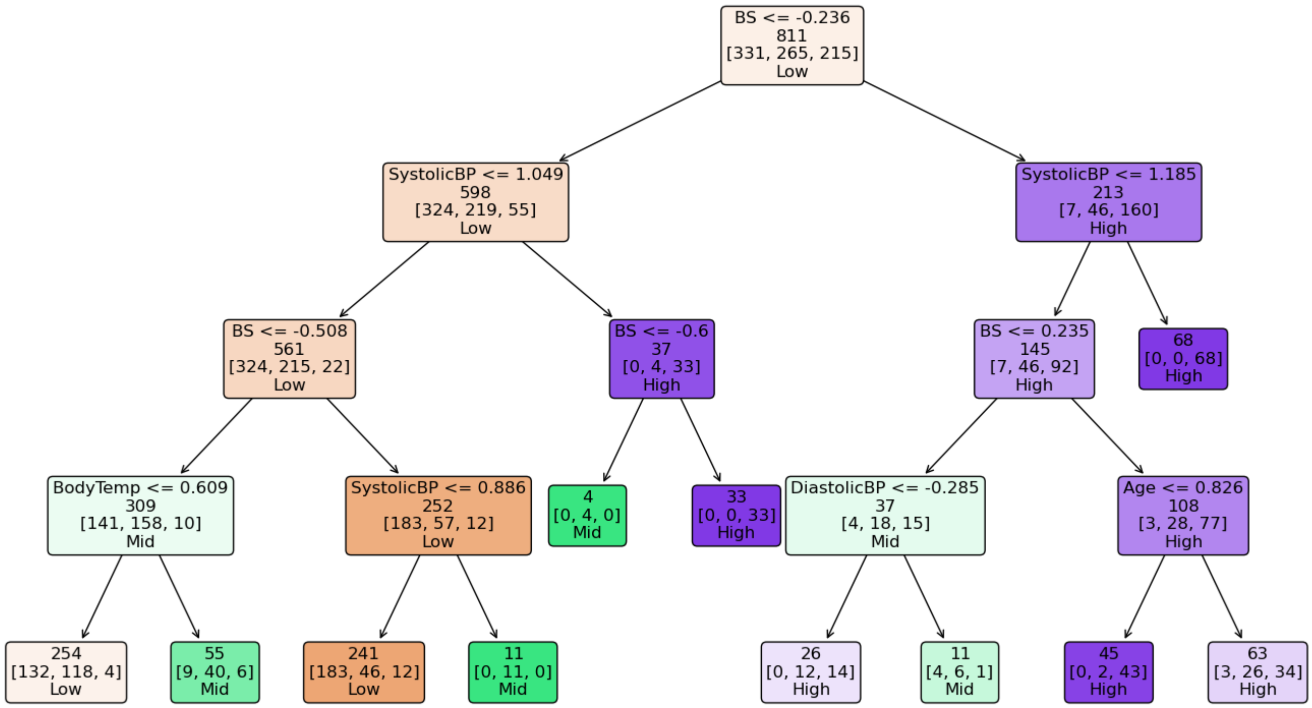

Figure 1: Decision tree diagram generated by decision tree model set to a maximum depth of 5 based on hyperparameter tuning. Each node displays the split criterion, the number of samples, class distribution [low, mid, high], and predicted risk class. Node colors represent each predicted class: orange for low, green for mid, and purple for high. Arrows show the decision made from node to node based on whether the split criterion is met (left for "yes", right for "no").

Table 2: The predictor variables present in the decision tree model with the corresponding importance, sorted in descending order.

Rank	**Feature**	**Importance**
1	Blood Glucose Level	0.635239
2	Systolic Blood Pressure	0.0277162
3	Body Temperature	0.040435
4	Age	0.035685
5	Diastolic Blood Pressure	0.01148
6	Heart Rate	0

Table 3 presents the predictor variables included in the random forest model, along with their corresponding importance. Importance was computed as the mean and standard deviation of the accumulation of the impurity decrease in each tree. The features were ranked in descending order based on their importance within the table. In terms of patterns, blood glucose level was the most important feature for classification, with an importance score of 0.357. Age and systolic blood pressure followed, with scores of 0.181 and 0.152, respectively. The remaining features—diastolic blood pressure, heart rate, and body temperature—had lower importance values, with body temperature ranked last at 0.054.

Table 3: The predictor variables present in the random forest model with the corresponding importance, sorted in descending order.

Rank	**Feature**	**Importance**
1	Blood Glucose Level	0.357474
2	Age	0.181074
3	Systolic Blood Pressure	0.151539
4	Diastolic Blood Pressure	0.138256
5	Heart Rate	0.117213
6	Body Temperature	0.054445

Figure 2 presents a heatmap of feature importance ranks for each of the three models. Based on this heatmap, blood glucose level (BS) consistently ranked as the most important feature across all models. Systolic blood pressure (SystolicBP) ranked moderately low in all three models, indicating more importance. Body temperature ranked in the moderately low range for both the ordinal logistic regression and decision tree models but ranked highest in the random forest model. Age was inconsistently ranked across the three models. Diastolic blood pressure (DiastolicBP) ranked moderately high in all three models, indicating less importance. Heart rate had a moderately high or the highest rank in all three models, indicating that it's the least importance feature.

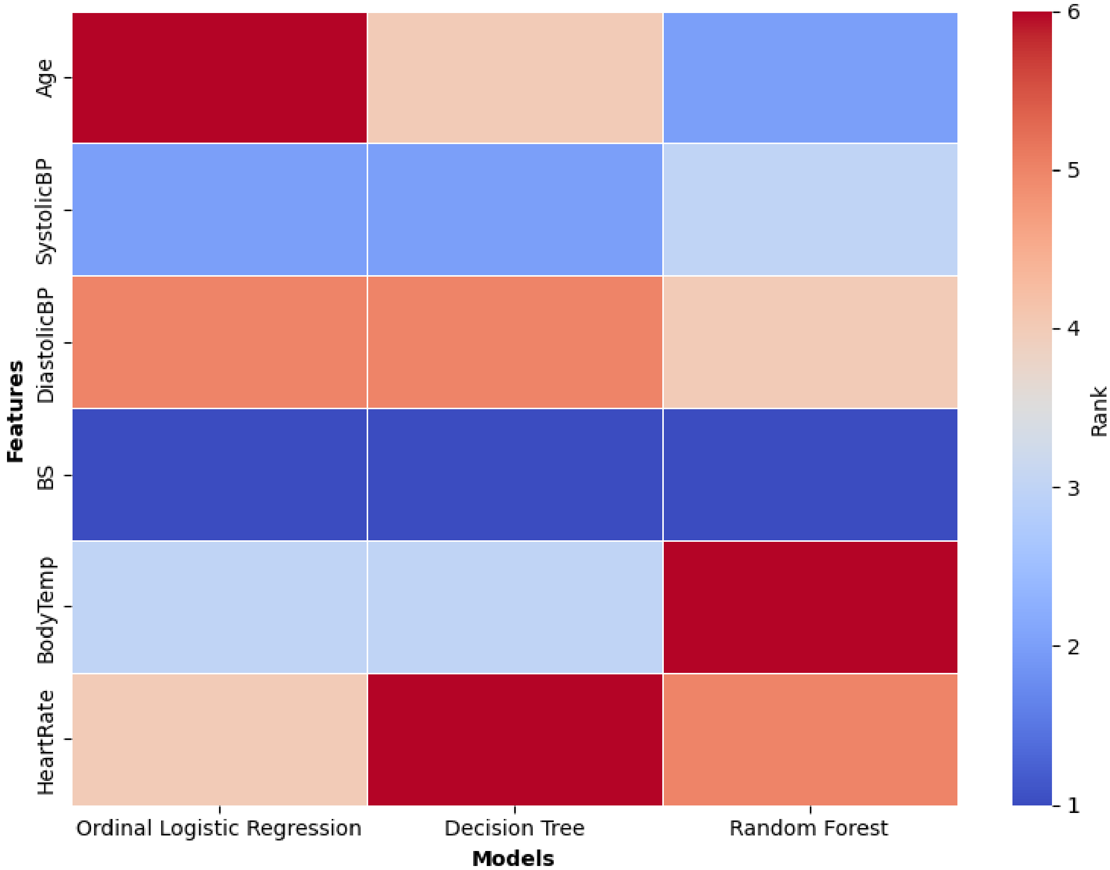

Figure 2: Heatmap showing feature importance rankings for each of the three models: ordinal logistic regression, decision tree, and random forest. Lower rank values indicate greater importance (dark blue), while higher rank values indicate lower importance (dark red).

Model Comparison

Table 4 presents a comparison of model accuracies based on their overall prediction performance. Accuracy was calculated as the sum of correct classifications divided by the total number of predictions. It includes three models—ordinal probit logistic regression, decision tree, and random forest—along with their corresponding accuracy percentages. The random forest model achieved the highest accuracy at 84.24%, followed by the decision tree model with 64.53%. The logistic regression model had the lowest accuracy, with a performance of 58.13%.

Figure 3 presents the confusion matrix for the ordinal logistic regression model. The model performed best on the low risk class, correctly predicting 60 out of 75 instances. It

Table 4: Model accuracy comparison based on the overall prediction accuracy for each model.

Model	**Accuracy (%)**
Random Forest	84.24
Decision Tree	64.53
Logistic Regression	58.13

did not perform as well for mid and high risk. For mid risk, the model had only 24 correct predictions out of 71 instances and for high risk, it correctly predicted 34 out of 57 instances. A large portion of mid risk cases (38) were misclassified as low risk.

Figure 4 presents the confusion matrix for the decision tree model. The model performed best on low risk class, correctly predicting 72 out of 75 instances. It also performed well on the high risk class, correctly predicting 44 out of 57 instances. However, for mid risk, the model had only 15 correct predictions out of 71 instances. A large portion of mid risk cases (47) were misclassified as low risk.

Figure 5 presents the confusion matrix for the random forest model. The model performed well for all three risk classes. For low risk, the model correctly predicted 59 out of 75 instances. For mid risk, the model correctly predicted 57 out of 71 instances. For high risk, the model correctly predicted 52 out of 57 instances.

Figure 6 presents the combined confusion matrix with aggregated classification results from all three models. Overall, the low and high risk classes were classified most accurately. For low risk, there were 191 correct predictions out of 225 instances and for high risk, there were 130 correct predictions out of 171 instances. However, the mid risk category had only 96 correct predictions out of 213. Notably, 97 mid risk instances were misclassified as low, and 20 as high.

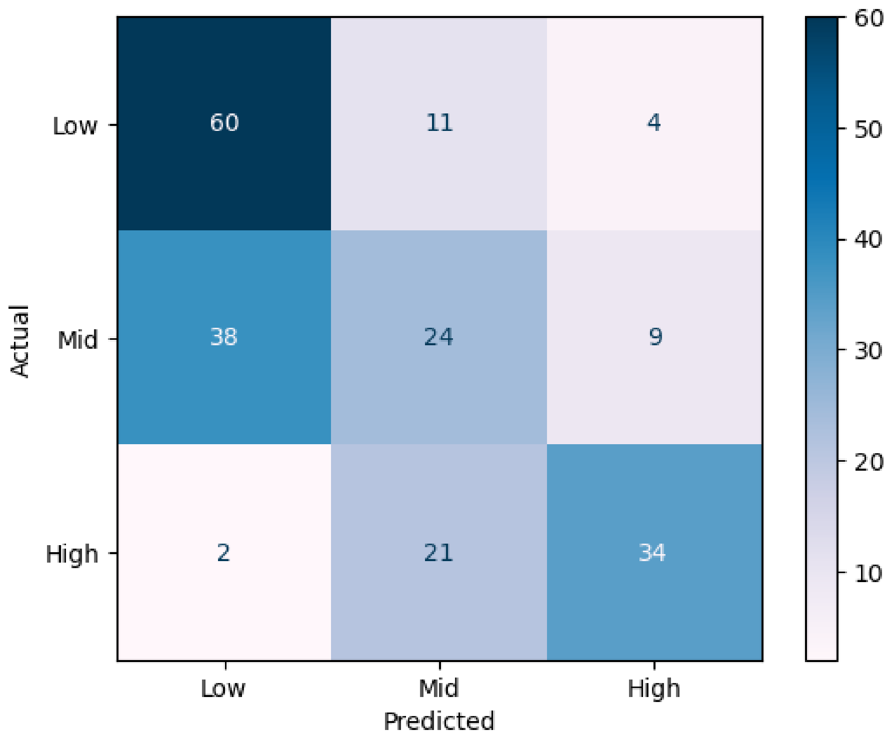

Figure 3: Confusion matrix for the ordinal logistic regression model to predict level of maternal health risk. It provided evaluation of its predictive performance in classifying maternal health risk into three categories: low, mid, and high. Each row of the matrix represents the true class, while each column represents the predicted class. Darker blue tones indicate more instances classified in that cell. The diagonal elements indicate the number of correct classifications for each class, while the off-diagonal elements represent misclassifications.

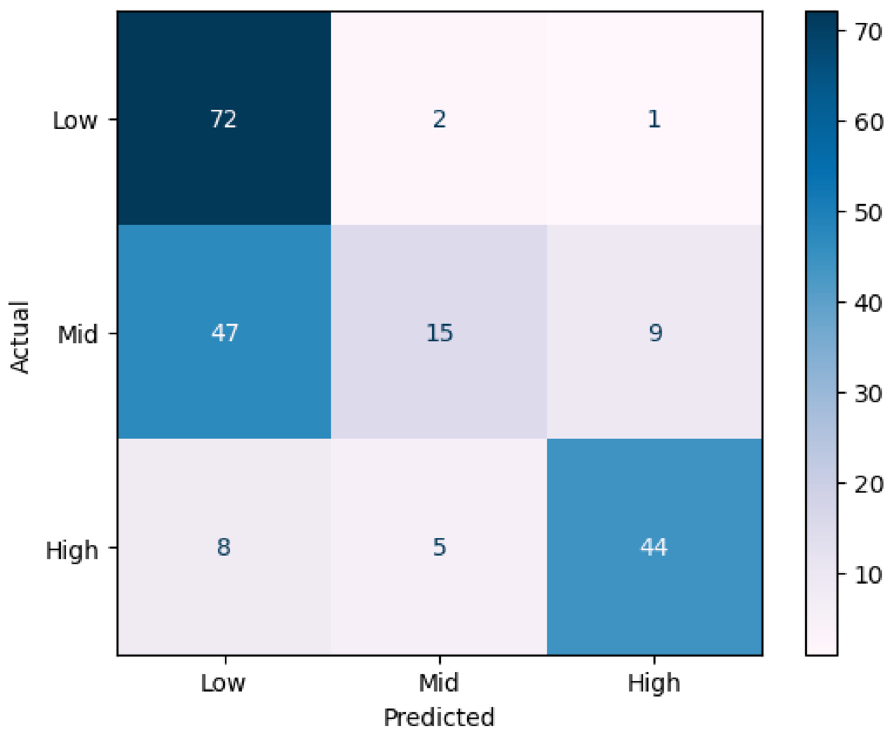

Figure 4: Confusion matrix for the decision tree fit with optimal maximum depth to predict level of maternal health risk. It provides evaluation of its predictive performance in classifying maternal health risk into three categories: low, mid, and high. Each row of the matrix represents the true class, while each column represents the predicted class. Darker blue tones indicate more instances classified in that cell. The diagonal elements indicate the number of correct classifications for each class, while the off-diagonal elements represent misclassifications.

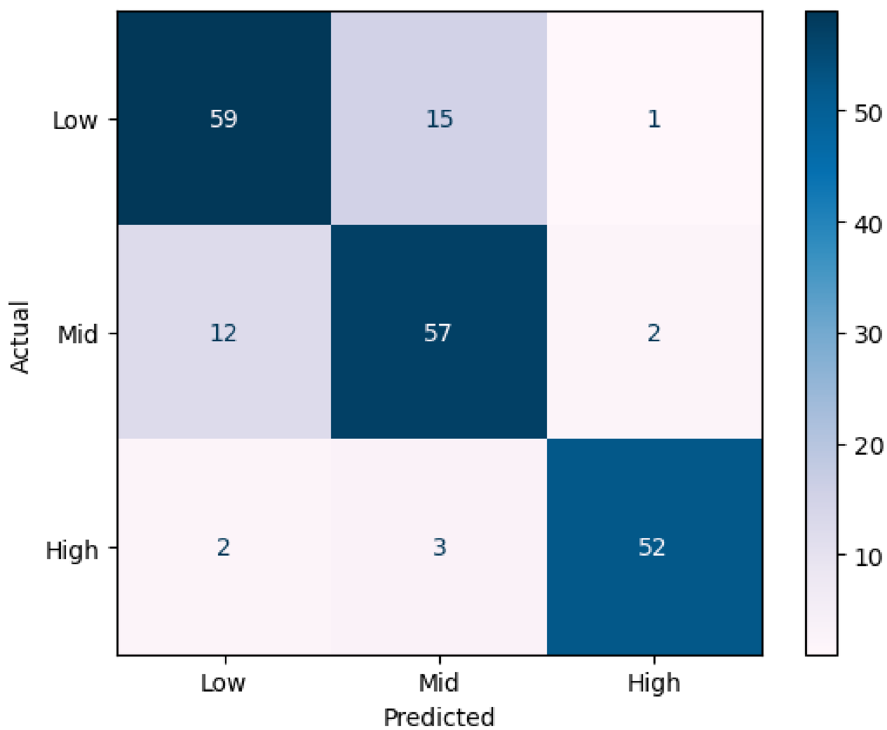

Figure 5: Confusion matrix for the random forest model fit with optimal maximum features to predict level of maternal health risk. It provides evaluation of its predictive performance in classifying maternal health risk into three categories: low, mid, and high. Each row of the matrix represents the true class, while each column represents the predicted class. Darker blue tones indicate more instances classified in that cell. The diagonal elements indicate the number of correct classifications for each class, while the off-diagonal elements represent misclassifications.

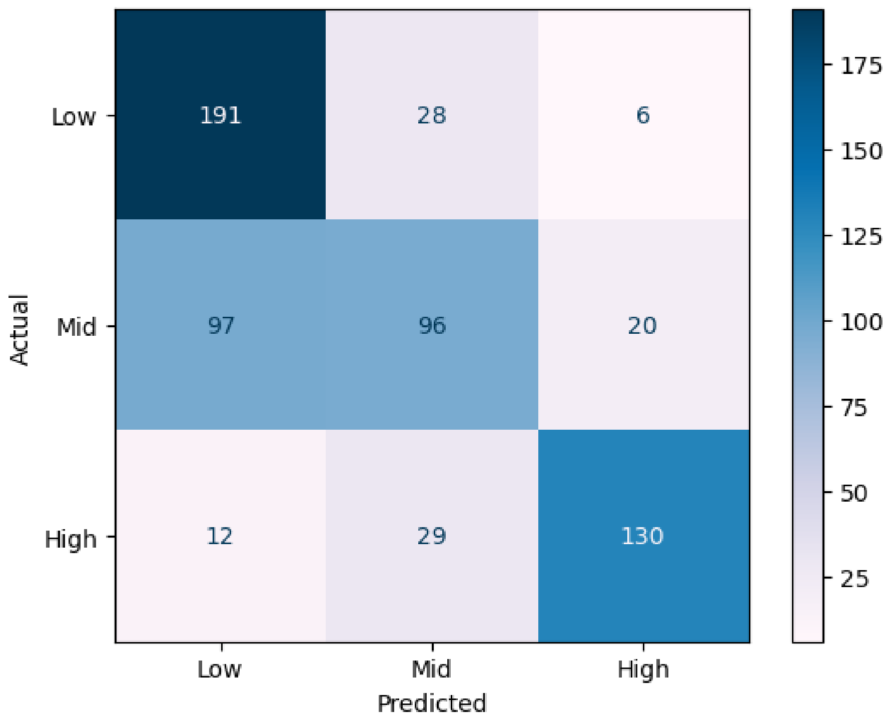

Figure 6: Combined confusion matrix showing the aggregated classification results from all three models: random forest, decision tree, and ordinal logistic regression. It provides evaluation of its predictive performance in classifying maternal health risk into three categories: low, mid, and high. Each row of the matrix represents the true class, while each column represents the predicted class. Darker blue tones indicate more instances classified in that cell. The diagonal elements indicate the number of correct classifications for each class, while the off-diagonal elements represent misclassifications.

Discussion

The research goals of this study were to identify the most important predictors for maternal health risk classification and to evaluate the performance of three ML models based on their overall accuracy and within-group accuracy.

In all three models, blood glucose level was ranked as the most important feature as shown in Figure 2. This aligns with previous research on maternal health outcomes. Multiple studies have consistently concluded that blood glucose levels affect maternal and infant health outcomes (Dodd et al., 2007; Zhao et al., 2023). The consistent high importance of blood glucose level across multiple models and studies suggests that it plays a critical role in assessing maternal health risk.

Systolic blood pressure was ranked moderately important in all three models. It indicates that it does seem to contribute to maternal health risk, however, it does not have the same predictive power as blood glucose level. In support of this finding, there are mixed findings about whether or not systolic blood pressure alone affects maternal health outcomes. Teng et al. (2021) concluded that patterns of high systolic blood pressure during pregnancy increase risk of poor maternal and fetal outcomes. However, another study concluded that other preexisting health factors affected health risk (Salles et al., 2014).

Moving onto the next feature, body temperature's importance ranking was very inconsistent across the three models. It was a moderately important feature in the ordinal logistic regression and decision tree models but least important in the random forest model.

Similarly, the importance of age as a predictor was also inconsistent among the three models. It was ranked moderately high in importance in the random forest model but ranked very low in importance in the ordinal logistic regression and decision tree models. This is an interesting finding because age is a commonly known risk factor for pregnancy (Sheen et al., 2018). One possible explanation is that the dataset's age range is limited. It contained an average age of 29 years old. The lack of variability in age could limit the simpler models' ability to find patterns related to age and risk level. Additionally, the ordinal logistic regression assumes a linear relationship between the predictor and outcome. Since logistic regression assumes a linear relationship, it may undervalue age as a predictor if its true relationship with risk level is nonlinear.

Diastolic blood pressure consistently ranked low in terms of importance in all three models. This aligns with a study by Steer et al. (2004), which found that diastolic blood pressure by itself cannot determine health risk. Specifically, they found that both low and high ranges of diastolic blood pressure increased risk of adverse health outcomes such as premature infants and maternal mortality (Steer et al., 2004).

So, while blood glucose level is consistently important for classifying maternal health risk across the board, other predictors vary across models. This shows that multiple models are needed to see the full picture.

Table 4 shows a substantial difference in classification performance between the models, with the random forest model outperforming the other two models by a considerable margin in terms of accuracy. Since a random forest is an ensemble method that combines predictions from multiple trees, it makes sense why it would outperform a decision tree and ordinal logistic regression model. As expected, the model with multiple trees predicted better than a single tree. The decision tree model performed slightly better than the ordinal logistic regression model. Based on this result, it suggests that the relationship between features and maternal health risk is nonlinear given that ordinal logistic regression assumes linearity.

For the ordinal logistic regression model, the confusion matrix (Figure 3) shows that the model reliably identifies low risk individuals and it struggles to accurately classify mid and high risk cases. These misclassifications reflect the assumption of linearity that the ordinal logistic regression model makes. It likely oversimplifies the nonlinear relationship between the features and the risk level. Using this model in applied settings would lead to the under-identification of higher risk patients. This is a concern that shows the limitations of simpler statistical models in high-stakes health contexts.

For the decision tree model, the confusion matrix (Figure 4) shows that while the model reliably identifies low and high risk individuals, it struggles to accurately classify mid risk cases. True mid risk classes were mostly classified as low risk. Again, this is a concern in application since it underestimated risk level.

For the random forest model, the confusion matrix (Figure 5) shows that the model can identify all three classes well.

The aggregated confusion matrix in Figure 6, presents the overall within-group classification performance of the combination of models. Misclassifications tend to lean toward low rather than high risk for mid risk instances. From a healthcare perspective, this could be a concern since the models underestimate risk level. On the other hand, the high-risk category has a low false positive rate, which is beneficial in practice as it reduces unnecessary medical attention for individuals who are not actually at high risk. This finding aligns with Liu et al. (2025), who reported a similar pattern in classifying mild cognitive impairment, noting that it helped reduce unnecessary burden.

Although, these results offer interesting insights into the important predictors and optimal models for classifying maternal health risk, there are several limitations to this study. First, the dataset lacks geographic diversity. All the data was collected from a single country, Bangladesh. This suggests that the results may not be generalizable to the whole population. In addition, the dataset also lacked socioeconomic and psychological variables that could be critical in predicting health risk. Socioeconomic features, such as education and poverty status, are consistently associated with maternal health outcomes (Gazmararian et al., 1996). Their absence in the dataset limits the models' ability to fully capture all predictors of risk. In terms of the models, the random forest model did outperform the other two ML models, but it only achieved an accuracy of 84.24% (Table 4). Since maternal health risk is a high-stake domain, small amounts of error can have significant consequences.

In the future, there are several different ways I could extend my research. One step that could be taken to improve this research would be to experiment with more complex models and tune additional hyperparameters. This could potentially capture more complex patterns in the data. To generalize these findings, I could evaluate the same models but incorporate a dataset with data from different regions. Additionally, I would like to incorporate demographic and socioeconomic data. This would expand my research on important predictors for classifying risk and also help identify risk among different patient groups. Finally, integrating information from Internet of Things (IoT) devices, such as data from wearable devices, would allow me to explore real-time risk evaluation.

References

Ahmed, M. (2020). Maternal Health Risk. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5DP5D.

Ahmed, M., Kashem, M. A., Rahman, M., and Khatun, S. (2020). Review and analysis of risk factor of maternal health in remote area using the internet of things (iot). *Springer*, pages 357–365.

Dodd, J. M., Crowther, C. A., Antoniou, G., Baghurst, P., and Robinson, J. S. (2007). Screening for gestational diabetes: The effect of varying blood glucose definitions in the prediction of adverse maternal and infant health outcomes. *Australian and New Zealand Journal of Obstetrics and Gynaecology*, 47(4):307–312.

Gazmararian, J. A., Adams, M. M., and Pamuk, E. R. (1996). Associations between measures of socioeconomic status and maternal health behavior. *American Journal of Preventive Medicine*, 12(2):108–115.

Goecks, J., Jalili, V., Heiser, L. M., and Gray, J. W. (2020). How machine learning will transform biomedicine. *Cell*, 181(1):92–101.

Liu, X., Ni, J., Wang, B., Yin, R., Tang, J., Chu, Q., You, L., Wu, Z., Cao, Y., and Ji, C. (2025). A prediction model for the risk of developing mild cognitive impairment in older adults with sarcopenia: evidence from the charls. *Aging Clinical and Experimental Research*, 37(1).

Macrohon, J. J. E., Villavicencio, C. N., Inbaraj, X. A., and Jeng, J.-H. (2022). A semi-supervised machine learning approach in predicting high-risk pregnancies in the philippines. *Diagnostics*, 12(11).

Mondal, S., Nag, A., Barman, A. K., and Karmakar, M. (2023). Machine learning-based maternal health risk prediction model for iomt framework. *International Journal of Experimental Research and Review*, 32:145–159.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Salles, G. F., Schlüssel, M. M., Farias, D. R., Franco-Sena, A. B., Rebelo, F., Lacerda, E. M., and Kac, G. (2014). Blood pressure in healthy pregnancy and factors associated with no mid-trimester blood pressure drop: A prospective cohort study. *American Journal of Hypertension*, 28(5):680–689.

Sathasivam, S. and Abdullahi, S. (2024). An ensemble learning algorithm with hyperparameter optimization (elaho) model for early detection of maternal health risk (mhr) level using machine learning. *Applied Mathematics and Computational Intelligence (AMCI)*, 13(3):66–86.

Seabold, S. and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.

Sheen, J.-J., Wright, J. D., Goffman, D., Kern-Goldberger, A. R., Booker, W., Siddiq, Z., D'Alton, M. E., and Friedman, A. M. (2018). Maternal age and risk for adverse outcomes. *American Journal of Obstetrics and Gynecology*, 219(4):390.e1–390.e15.

Steer, P. J., Little, M. P., Kold-Jensen, T., Chapple, J., and Elliott, P. (2004). Maternal blood pressure in pregnancy, birth weight, and perinatal mortality in first births: prospective study. *BMJ*, 329(7478):1312.

Teng, H., Wang, Y., Han, B., Liu, J., Cao, Y., Wang, J., Zhu, X., Fu, J., Ling, Q., Xiao, C., Wan, Z., and Yin, J. (2021). Association between maternal blood glucose levels during pregnancy and birth outcomes: A birth cohort study. *BMC Pregnancy Childbirth*, 21.

Togunwa, T. O., Babatunde, A. O., and Abdullah, K.-u.-R. (2023). Deep hybrid model for maternal health risk classification in pregnancy: synergy of ann and random forest. *Frontiers in Artificial Intelligence*, 6.

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The cancer genome atlas (tcga): An immeasurable source of knowledge. *Wspolczesna Onkologia*, 1A:A68 – A77. Cited by: 1656; All Open Access, Gold Open Access, Green Open Access.

Venter, J. (2001). The sequence of the human genome. *Science*, 291(5507):1304 – 1351. Cited by: 81.

World Health Organization (2024). Maternal mortality.

Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., Stein, L. D., and Ferretti, V. (2019). The international cancer genome consortium data portal. *Nature Biotechnology*, 37:367–369.

Zhao, D., Liu, D., Shi, W., Shan, L., Yue, W., Qu, P., Yin, C., and Mi, Y. (2023). Association between maternal blood glucose levels during pregnancy and birth outcomes: A birth cohort study. *International Journal of Environmental Research and Public Health*, 20(3).